5. Multivariate Differentiation

5.1. Total Differentiation and The Chain Rule

Definition 1. Suppose E is an open set in R^n, f maps E into R^m, and $\mathbf{x} \in E$. If there exists a linear transformation A of R^n into R^m such that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}|}{|\mathbf{h}|} = 0, \tag{1}$$

then we say that $\mathbf{f}$ is *differentiable* at $\mathbf{x}$, and A would be unique. We write

$$\mathbf{f}'(\mathbf{x}) = A.$$

If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say that $\mathbf{f}$ is *differentiable* in E.

Remark. (a) It is of course understood that $\mathbf{h} \in R^n$, and $|\cdot|$ denotes the norm in R^n. Also, it is better to regard the linear transformation A as a $m \times n$ matrix.
(b) The relation (1) can be rewritten in the form

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h}) \tag{2}$$

where the remainder $\mathbf{r}(\mathbf{h})$ satisfies

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0.$$

Thus, for fixed $\mathbf{x}$ and small $\mathbf{h}$, the left side of (2) is approximately equal to $\mathbf{f}'(\mathbf{x})\mathbf{h}$, that is, to the value of a *linear* transformation applied to $\mathbf{h}$.
(c) A glance at (2) shows that $\mathbf{f}$ is continuous at any point at which $\mathbf{f}$ is differentiable.
(d) The derivative defined by (1) is often called the *differential* of $\mathbf{f}$ at $\mathbf{x}$, or the *total derivative* of $\mathbf{f}$ at $\mathbf{x}$.

Example 1. For a linear transformation $A \in \mathscr{L}(R^n, R^m)$ and $\mathbf{x} \in R^n$, we have

$$A'(\mathbf{x}) = A$$

The proof is a triviality, since

$$A(\mathbf{x}+\mathbf{h}) - A(\mathbf{x}) = A\mathbf{h},$$

by the linearity of A.

Theorem 1. (Chain rule). Suppose E is an open set in R^n, $\mathbf{f}$ maps E into R^m, $\mathbf{f}$ is differentiable at $\mathbf{x}_0 \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ into R^k, and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x}_0)$. Then the mapping $\mathbf{F}$ of E into R^k defined by

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x}_0$, and

$$\mathbf{F}'(\mathbf{x}_0) = \mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0). \tag{3}$$

On the right side of (3), we have the product of two linear transformations.

5.2. Partial Differentiation, Jacobian and Gradient

Definition 2. Consider a function $\mathbf{f}$ that maps an open set $E \subset R^n$ into R^m. Let $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ be the standard bases of R^n and R^m. The *components* of $\mathbf{f}$ are the real functions $f_1, \cdots, f_m$ defined by

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{u}_i \quad (\mathbf{x} \in E),$$

or, equivalently, by $f_i(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{u}_i, 1 \le i \le m$. For $\mathbf{x} \in E, 1 \le i \le m, 1 \le j \le n$, we define

$$(D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t},$$

provided the limit exists. Writing $f_i(x_1, \cdots, x_n)$ in place of $f_i(\mathbf{x})$, we see that $D_j f_i$ is the derivative of f_i with respect to x_j, keeping the other variables fixed. The notation

$$\frac{\partial f_i}{\partial x_j}$$

is therefore often used in place of $D_j f_i$, and $D_j f_i$ is called a *partial derivative*.

Remark. If $\mathbf{f}$ is known to be differentiable at a point $\mathbf{x}$, then its partial derivatives exist at $\mathbf{x}$, and they determine the linear transformation $\mathbf{f}'(\mathbf{x})$ completely:

Let $[\mathbf{f}'(\mathbf{x})]$ be the matrix that represents $\mathbf{f}'(\mathbf{x})$ with respect to our standard bases, then $\mathbf{f}'(\mathbf{x})\mathbf{e}_j$ is the jth column vector of $[\mathbf{f}'(\mathbf{x})]$, and the number $(D_j f_i)(\mathbf{x})$ occupies that spot in the ith row and jth column of $[\mathbf{f}'(\mathbf{x})]$. Thus

$$[\mathbf{f}'(\mathbf{x})] = \begin{pmatrix} (D_1 f_1)(\mathbf{x}) & \cdots & (D_n f_1)(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ (D_1 f_m)(\mathbf{x}) & \cdots & (D_n f_m)(\mathbf{x}) \end{pmatrix}$$

This matrix is called the *Jacobian matrix*, denoted by $D\mathbf{f}, \mathbf{J_f}, and \frac{\partial(f_1,\cdots,f_m)}{\partial(x_1,\cdots,x_n)}$.
If m=n, the Jacobian matrix is a square matrix, and its determinant is the *Jacobian determinant* of **f**. In this case, the matrix and its determinant are both referred to as the *Jacobian* in literature.
If $m = 1$, the Jacobian matrix is reduced to a row vector of partial derivatives of **f**, that is, the *gradient* of **f**, denoted by $\Delta\mathbf{f}$.

Example 2. If $f(x,y) = x^y$, then

$$\frac{\partial f}{\partial x} = yx^{y-1}, \frac{\partial f}{\partial y} = x^y \ln x.$$

The chain rule naturally applies to partial derivatives. If $z = z(y)$ is a mapping from R^n to R and $y = y(x)$ is a mapping from R^m to R^n, then

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial z}{\partial y_j}\frac{\partial y_j}{\partial x_i}.$$

Exercise 3. Let $f(x,y) = x\ln(x^2+y^2)$. Calculate its partial derivatives.

Exercise 4. Let $f(x,y,z) = (x^2+y^2)z^2 + \sin x^2$. Calculate its partial derivatives.

Exercise 5. Let $z = z(u,v) = v\ln u, u = x^2+y^2, v = \frac{y}{x}$. Calculate $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

Exercise 6. Let $z = e^{xy}\sin(x+y)$. Calculate $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$.

Example 7. Let $f(x,y,z) = \left(\frac{y}{x}\right)^z$. Then

$$(\Delta f)(x,y,z) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) = \left(-\frac{zy^z}{x^{z+1}}, \frac{zy^{z-1}}{x^z}, \left(\frac{y}{x}\right)^z \ln\left(\frac{y}{z}\right)\right).$$

Exercise 8. Let $f(x,y) = x^2+2xy+y^2$. Calculate $(\Delta f)(1,2)$.

Exercise 9. Let $F(x,y) = (x^2+y^3, xy)$. Calculate the Jacobian J_F.

5.3. Continuous Differentiability

Definition 3. Let $T \in \mathscr{L}(R^n, R^m)$. Define the norm $||T||$ of T to be the supremum of all numbers $|T(\mathbf{x})|$, where $\mathbf{x}$ ranges over all vectors in R^n with $|\mathbf{x}| \leq 1$.

Theorem 2. If $T, U \in \mathscr{L}(R^n, R^m)$ and c is a scalar, then

$$||T+U|| \leq ||T|| + ||U||, \quad ||cT|| = |c| \cdot ||T||.$$

With the distance between T and U defined as $||T-U||$, $\mathscr{L}(R^n, R^m)$ is a metric space.

Definition 4. A differentiable mapping $\mathbf{f}$ of an open set $E \subset R^n$ into R^m is said to be *continuously differentiable* in E if $\mathbf{f}'$ is a continuous mapping of E into $\mathscr{L}(R^n, R^m)$.
More explicitly, it is required that to every $\mathbf{x} \in E$ and to every $\varepsilon > 0$ corresponds a $\delta > 0$ such that

$$||\mathbf{f}'(\mathbf{x}) - \mathbf{f}'(\mathbf{y})|| < \varepsilon$$

if $\mathbf{y} \in E$ and $|\mathbf{x} - \mathbf{y}| < \delta$. If this is so, we also say that $\mathbf{f}$ is a $\mathscr{C}'$-mapping, or that $\mathbf{f} \in \mathscr{C}'(E)$.

Theorem 3. Suppose $\mathbf{f}$ maps an open set $E \subset R^n$ into R^m. Then $\mathbf{f} \in \mathscr{C}'(E)$ if and only if the partial derivatives $D_j f_i$ exists and are continuous on E for $1 \leq i \leq m, 1 \leq j \leq n$.

5.4. The Inverse Function Theorem

The inverse function theorem states, roughly speaking, that a continuously differentiable mapping $\mathbf{f}$ is invertible in a neighborhood of any point $\mathbf{x}$ at which the linear transformation $\mathbf{f}'(\mathbf{x})$ is invertible and also gives the expression for the derivative of the inverse function.

Theorem 4. Suppose $\mathbf{f}$ is a $\mathscr{C}'$-mapping of an open set $E \in R^n$ into R^n, $\mathbf{f}'(\mathbf{a})$ is invertible for some $\mathbf{a} \in E$, and $\mathbf{b} = \mathbf{f}(\mathbf{a})$. Then
(a) there exists open sets U and V in R^n such that $\mathbf{a} \in U, \mathbf{b} \in V$, $\mathbf{f}$ is one-to-one on U, and $\mathbf{f}(U) = V$;
(b) if $\mathbf{g}$ is the inverse of $\mathbf{f}$, defined in V by

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x} \quad (\mathbf{x} \in U),$$

then $\mathbf{g} \in \mathscr{C}'$ and

$$\mathbf{g}'(\mathbf{y}) = [\mathbf{f}'(\mathbf{g}(\mathbf{y}))]^{-1} \quad (\mathbf{y} \in V).$$

Example 10. Let $g(x) = x^n$ for $x > 0$. Let f be the inverse function of g. Then

$$f'(x) = \frac{1}{g'(f(x))} = \frac{1}{n(x^{1/n})^{n-1}} = \frac{1}{n}x^{1/n-1}.$$

Exercise 11. Prove $(\ln x)' = \frac{1}{x}, x > 0$, provided $(e^y)' = e^y$.